

IoT 환경에서 스트리밍 기반의 비정형 데이터 수집 프레임워크 설계

이후영⁰, 박구락^{*}, 김동현^{**}

⁰공주대학교 멀티미디어공학과

^{*}공주대학교 컴퓨터공학부

^{**}우송대학교 IT융합학부

e-mail: hooyoung.paul.lee@gmail.com⁰, ecgrpark@kongju.ac.kr^{*}, dhkim@wsu.ac.kr^{**}

Design of Streaming based Unstructured-Data Collecting Framework in IoT Environment

Hoo-Young Lee⁰, Koo-Rack Park^{*}, Dong-Hyun Kim^{**}

⁰Dept. of Multimedia Engineering, Kongju National University

^{*}Dept. of Computer Science & Engineering, Kongju National University

^{**}Dept. of IT Convergence, Woosong University

● 요약 ●

사물인터넷 환경의 다양한 기기에서는 매초마다 시스템 로그 데이터, 온도, 습도, 조도 및 위치 정보 등과 같은 데이터를 지속적으로 생성한다. 이렇게 생성된 데이터는 기기 안에서 대부분 소멸되거나 수집된다 하더라도 시스템 개선의 일부 목적으로 활용하는데 그칠 뿐이다. 본 논문에서는 각각의 사물인터넷 기기에서 발생하는 비정형 데이터를 스트리밍 방식을 통해 수집 서버로 전송하고 이를 유연한 스키마 구조를 가지는 NoSQL 데이터베이스에 적재하는 프레임워크 설계를 제안한다. 이렇게 수많은 장비로부터 수집된 로그 및 센싱 데이터는 빅데이터 분석을 통해 산업의 현장에서 생산성 향상을 위해 사용할 수 있으며 공공의 목적으로 도심지의 교통문제 해소와 재난 및 재해 예측에 활용될 수 있다.

키워드: 사물인터넷(Internet of Things), 비정형데이터(Unstructured Data), NoSQL

I. Introduction

사물인터넷(Internet of Things, IoT)은 현대 무선 통신에서 빠르게 발전하고 있는 새로운 패러다임이다. 세계적인 컨설팅 기업 중 하나인 맥킨지는 2025년 까지 인류의 삶을 가장 급진적으로 변화시킬 기술 중 하나로 사물인터넷 기술을 선정하고, 향후 거의 모든 산업 분야에서 사물 인터넷 기술이 적용될 것으로 전망했다[1].

이러한 사물인터넷 환경에서 각각의 디바이스에서 생성되는 다양한 형태의 비정형 로그 데이터는 대부분 소실되거나 장치(Device) 자체의 오류 개선을 위해 사용되는 정도로만 활용될 뿐 거시적 관점에서 데이터를 수집/분석하는 작업이 이뤄지지

않고 있다. 그러나 각종 센서에서 발생하는 다양한 데이터를 수집할 수 있도록 장치들 간에 표준 데이터 수집 프레임워크를 이식함으로써 소실되는 데이터를 스트리밍 방식을 통해 단일 수집 서버에 저장할 수 있고, 저장된 대량의 비정형 데이터는 빅데이터 분석을 통해 산업현장에서 생산력을 높이거나, 공공 부분의 재난 예측 및 재해 방지 등 사회 전반에 걸쳐 다양한 분야에 활용될 수 있다.

본 논문에서는 클라이언트/서버 방식을 통해 각각 장치로부터 텍스트 기반의 데이터를 수집하고 이를 실시간으로 비정형 데이터 처리에 적합한 NoSQL 서버에 저장하는 프레임워크를 제안한다.

II. Related Works

1. 사물인터넷

사물인터넷(IoT : Internet of Things)은 각각의 사물에 통신이 가능한 센서를 장착하여 인터넷에 연결하는 기술을 의미한다. 여기서 사물에 대한 정의는 생산 현장의 각종 장비나 가정에서 사용하는 여러가지 전자제품 혹은 웨어러블 컴퓨터(Wearable Computer) 등을 의미한다. 사물인터넷은 감각, 통신, 네트워킹 및 정보처리기술을 기반으로 하는 수많은 장치들의 연결로 구성된 글로벌 네트워크 인프라라고 간략히 정의 할 수 있다[2].

3. NoSQL

NoSQL(Not only SQL)은 비관계형 데이터베이스이다. 미리 정의된 스키마를 가지고 있는 관계형 데이터베이스는 많은 분야에서 사용되어 왔지만 비정형 데이터를 처리함에 있어서 한계가 있다. NoSQL 데이터베이스는 유연한 스키마 구조와 데이터의 증가에 따른 노드 분산을 통하여 데이터베이스 구조를 쉽게 확장할 수 있기 때문에 많은 분산 시스템 개발자들에게 주목을 받고 있다[3][4].

III. The Proposed Scheme

제안 시스템은 사물인터넷 환경에서 다수의 장치(Device)에서 발생하는 비정형 데이터 로그를 스트리밍 방식으로 수집하여 데이터 분석 서버에 저장하는 프레임워크(Framework)에 대한 설계로서 크게 3가지 영역으로 구성되어 있다.

첫째, 데이터 수집 클라이언트(Data Collection Client) 부분으로, 클라이언트 프로그램은 각각의 IoT 장비에 설치되어 장치에서 실시간으로 발생하는 텍스트 기반의 센싱 데이터를 소켓 스트리밍 방식으로 서버에 전송하는 역할을 수행한다. 둘째, 데이터 수집 서버(Data Collection Server) 부분으로, 데이터 수집 서버는 IoT 장치에 설치된 클라이언트 프로그램과 통신하며 각각의 장치에서 전송하는 다양한 형태의 센싱 데이터를 수집하여 파일형태로 서버에 저장한다. 셋째, 전송된 파일에서 데이터를 파싱하여 NoSQL 기반의 데이터베이스에 저장하는 역할을 수행한다.

다음의 [Fig. 1]은 제안시스템의 구조도로써, IoT 장비로부터의 데이터 수집과 수집된 데이터의 저장, 수집된 다양한 형태의 데이터를 NoSQL 서버에 적재하는 과정을 나타낸다.

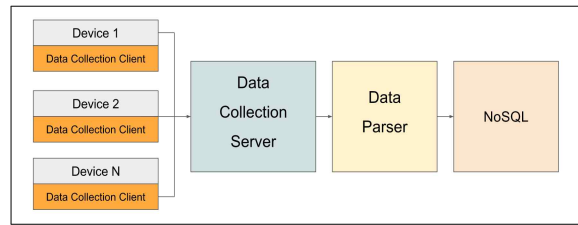


Fig. 1. System Architecture

IV. Conclusions

사물인터넷 환경에서 사용되는 장치들은 매초마다 다양한 형태의 데이터를 발생한다. 발생하는 로그의 내용은 보통 기기에서 발생하는 단순 로그에서부터 시작해서 온도, 습도, 조도, 위치 정보 등과 같은 비정형 데이터 들이다. 본 논문은 사물인터넷 환경에서 발생하는 다양한 형식의 텍스트 기반 로그데이터를 수집하여 이를 비정형 데이터 처리에 최적화되어 있는 NoSQL 서버에 저장하는 프레임워크에 대한 설계이다. 데이터 수집은 각 장비에 별도의 수집 프로그램을 설치하고 무선 통신을 통하여 수집 서버에 저장된다. 수집서버에 저장된 데이터는 별도의 데이터 파서를 통해 NoSQL 서버에 저장된다. 저장된 데이터를 이용하기 위하여 다양한 분석기법을 활용하면, 산업 현장에서의 생산력 증가하거나 공공 부문에 재난과 재해를 방지하는데 사용될 수 있을 것으로 기대된다.

References

- [1] James Manyika, Michael Chui, Jacques Bughin, and Richard Dobbs, Peter Bisson, and Alex Marrs, "Disruptive technologies: Advances that will transform life, business, and the global economy," McKinsey Global Institute, May 2013.
- [2] Y.J.Kim, H.Y.Yoon, "Service Oriented Architecture for Internet of Things", Conf. of KSCI, 2015.
- [3] Neal Leavitt, "Will NoSQL Databases Live Up to Their Promise?", Computer, Vol. 43, No. 2, pp.12-14, 2010. 02.
- [4] Jing Han, Haihong E., Guan Le, Jian Du, "Survey on NoSQL database", Pervasive Computing and Applications (ICPCA) 2011 6th International Conference on, pp.363-366, 2011. 10.